DOCUMENT RESUME

ED 167 887 CG 013 284

AUTHOR Lutkus, Anthony D.
TITLE Using PEERRATE: A Computerized System of Student Term Paper Grading.
PUB DATE 30 Aug 78
NOTE 22p.; Paper presented at the Annual Convention of the American Psychological Association (86th, Toronto, Ontario, Canada, August, 1978); Best copy available

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS College Students; *Computer Oriented Programs; *Evaluation Criteria; *Evaluation Methods; *Grade Equivalent Scores; Higher Education; *Peer Evaluation; *Performance Criteria; Rating Scales; Scoring Formulas; Writing
IDENTIFIERS *PEERRATE

ABSTRACT

The use of a computerized system is described in which students rate each other's term papers via the computer program PEERRATE, which creates a matrix of intervals between ratings received and calculates final ratings incorporating a mean and standard deviation selected by the instructor for the class. As described, the program includes: (1) compensation for the tendency of students to give high or low grades to everyone; (2) computation of a norm-adjusted score for performance by each student on each assignment; and (3) analysis of both the quality of the ratings by raters for that assignment and the consistency of each student as a rater. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED167887

USING PEERRATE: A COMPUTERIZED SYSTEM OF STUDENT TERM PAPER GRADING

A paper read at
American Psychological Association Convention
Toronto, Canada
August 30, 1978

BEST COPY AVAILABLE

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

A Lutkus

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Anthony D. Lutkus, Ph.D.
Office of the Dean
Rutgers University
325 Hill Hall
Newark, New Jersey 07102
(201) 648-5213

CG 013284

Can undergraduates with disparate backgrounds meaningfully grade each other's college assignments? This paper will describe a computer system developed at Rutgers University-Newark that facilitates student ratings of each other's work. A basic assumption underlying the development of this system is that students will have an enriched learning experience if they participate in specifying grading criteria and in evaluating student assignments according to those criteria.

This report is in three sections: First, a brief background of the PEER-RATE computer program; second, a consideration of how the system has been used in undergraduate courses; and third, preliminary results from PEERRATE and a summary of student reaction gleaned from surveys taken at the end of courses which employed the system.

BACKGROUND OF THE SYSTEM

Over the past few years, MBA students at Rutgers-Newark Graduate School of Business have been rating each other's performance on team projects for business corporations. Inherent in any rating task where the participants rate only a subset of the total group is the possibility of individual rater bias (either too lenient or too stiff) distorting the ratings. In addition, there is the human problem of students tending to want to give each other an "A" and, consequently, using only the high end of any rating scale. To solve these procedural difficulties, Dr Purnell Benson developed (1976, 1977) the PEERRATE program which takes student-provided ratings as input, reduces rater bias, adjusts the rating distribution to a pre-set mean and deviation, and then prints

a report for each participant including his calculated grade and his efficiency as a grader.

The most obvious solution to the rater bias problem would be to have raters rank all performers from first to last, thus avoiding the "all-A" problem. This means, however, that every student rater has to rate every performer -- an impractical situation in a large class. One can, however, ask raters to rate a subset of the class on a scale with behaviorally defined categories like "excellent", "very good", "fair", or "outrageous". The point intervals or steps between different performers can then be established. Now everyone need not rate everyone else. The size of the intervals between rated performers establish the differences in their achievement.

Even if a class of raters tends to use only the upper end of a rating scale, the intervals between performers remain. Thus, it is a simple algebraic manuever to spread out the part of the scale that has been used, maintaining the rank orders between performers but setting a new mean and deviation. The average score given by the raters sets the achievement norm for the entire group. An absolute measure of achievement can be designated by an outside authority (usually the teacher!) who knows what the group as a whole has produced.

The PEERRATE program solves the problem of incomplete sets of rating intervals by using Mosteller's technique (1951) of finding the best set of final ratings which agree most closely with the inputed rating data. It is a least squares procedure in which the least errors of inconsistency remain between the calculated difference separating performers and the original differences reported

by the raters.

Other procedural problems to be solved before peer rating could be made practical included: providing an evaluation of rating skill so students will take the rating task seriously, finding a way for the teacher to detect faulty or discrepant ratings so they can be adjusted, and, finally, providing anonymity so that raters cannot favor their friends.

Dr. Benson's PEERRATE program automatically outputs a measure of the rater's skill in rating. It is possible, then, for the instructor to assign a "grade for grading". Four separate but related measures of rating consistency are used. First, the individual grades given by each rater are compared with the grades given to the same performers by the class as a whole. The latter, general grades are given by computer calculation and are called "adjusted group ratings", or AGR's Thus, the AGR's provide one standard for evaluating raters. The instructor can substitute his own grading on the assignment for the AGR's if he wishes. In fact, the program can include the instructor's grades along with student-generated grades and weight each according to whatever proportion the instructor deems appropriate.

The second evaluation of rater skill addresses the average bias of each rater. Does he or she rate too high or too low in general? The computer reports a "bias" score for each rater. Thirdly, the range of the rating scale that the rater uses can provide an indicator of his rating skill. Does he use more or less of the scale than the entire class? Again, the computer produces a difference score between the individual rater's range and that of other raters.

Fourth, after adjusting the grades for systematic bias and for the range used, the computer prints the standard deviation of the residual error that remains between the rater's ratings and the AGR's. If students also serve as raters for their own work, the discrepancy between self-rating and class rating can also be computed. These measures of rating skill are combined by the computer into a single score for rating ability. This composite rating ability score could conceivably serve as one component in the student's course grade.

Aside from producing a measure of each rater's skill at rating, Benson's program makes it possible for the instructor to assess the quality of all the ratings that went into a single performer's Adjusted Group Rating. AGR's with low agreement can then be reviewed by the instructor and possibly changed. PEERRATE prints a table of rank ordered scores for ratee performance, quality of ratee scores, and rater ability. The instructor can review the individual work involved to see whether, in particular, low grades with low agreement are justified. If the grades with low agreement are reasonable, then the instructor can assume that the AGR's for the rest of the class are valid. The instructor can review, for instance, a sample of term paper assignments beginning with the low AGR's some middle level scores, and a few at the top of the distribution. If the papers seem out of line with the absolute score value that has been set for the class norm, the batch can be run again with a different norm. Alternatively, it may be that the class was either not properly prepared for the grading task or did not absorb sufficient course content.

USE OF PEERRATE AT RUTGERS-NEWARK

The PEERRATE program is adaptable to several types of class assignments. Written essays, oral reports, term reports, team reviews, or co-written tests can each be handled by the program. My own use of the system has focused on grading term papers in large classes of introductory and child psychology. It has been my experience that one can assign long term papers in large classes (normally taboo to an over-worked instructor) by using PEERRATE to assist in the grading process.

Previously mentioned problems of anonymity and avoidance of peer favoritism can be handled by classroom procedures that are almost as important as the computer program itself. Students hand in multiple copies of their assignment; one for each rater. Computer identification numbers are assigned to each student and all identifying information is then removed from the papers. In a subsequent class, each student receives a package of papers to grade (usually between 3 and 6), an answer sheet for the grades, and, in our instance, a questionnaire to complete regarding the system. Student numbers are assigned from 2 (number 1 being reserved for the instructor's grades, if used) to N+1, with N being the class size. The program is not affected by extra numbers or missing numbers because it automatically contracts the matrix size during the calculations. At present, 35 is the maximum number of raters that can be inputted. Larger classes can be served by splitting them and treating the subsections independently.

The answer or rating form is prepared by the instructor. Appendix A displays

a sample of one used in my courses. The student rater's name is included for ease of distribution but the ratee's names remain anonymous.

A feature that makes PEERRATE a useful pedagogical tool is the provision for multiple ratings, i.e., one for each of several criteria. I have found that it makes for a stimulating class exercise to have students develop the criteria themselves. When the assignments are term papers, the criteria selected have followed nearly the same pattern over four semesters: content, organization, grammar or style, originality, and overall worth. In addition, the relative weights of each of the criteria can be set for each run of the program. Students enjoy debating -- at length -- whether, for example, content and organization should be of equal weight with grammar, style, etc. From the instructor's point of view, such thought about the prerequisites of a good paper can only be beneficial. The instructor then explains his interpretation of the criteria, possibly gives examples and then includes brief written descriptions (see Appendix B) of the criteria at the time the papers are distributed for rating. The program can accept any range of scales so it is possible to use, say, a 0 to 20 scale instead of the standard 1 to 100. The purpose of this is to expand the range of the scale students tend to use. If 1 to 100 is used, students tend to think of 70 as "passing". The wider the spread of scale intervals employed by raters, the better will be the discrimination among ratees. On a scale of 1 to 20 or 1 to 35, the preconceived "anchor" of "70" is harder to define and the rater tends to rely on the category labels attached to the scale by the instructor.

When a given term paper is rated by less than the entire class, the

numerical assignment of ratees to each rater becomes extremely important. The subsequent paired-comparison calculation by PEERRATE assumes that there is, in Benson's words, "a distributed, balanced design". "Distributed" refers to the requirement that each student be rated by a different subgroup of peers, i.e., the combination of rater identification numbers is unique for each rater. "Balanced" means that each subgroup of raters is scattered through the class distribution, not bunched into clusters. The assignment of the subgroup of ratees for the first student rater (#2) would proceed as follows (Benson, 1976):

> "...Then students to be rated by the lowest rater, 2, are selected by taking subscript numbers which fall proportionately from the largest ID number, say 30. Using 2/3 as the proportion, the ID numbers selected for six ratees, rounding to the nearest integer would be 20, 14, 10, 7, 5, 3. This same configuration is extended to other raters by increments. If the rater's ID number is, say, 20, then add 18 to each in this configuration, yielding 9, 3, 28, 25, 23, and 21."

More detailed examples of ID number assignment procedures will be available in a PEERRATE handbook that is in preparation (Benson, 1978).

Once the multiple copies of student papers are received, the instructor or an assistant assigns the ID numbers to the papers and collates the package of papers to be returned to each rater at the next class meeting. Students can either do the rating in class or take the papers home for a more leisurely reading. The rating sheets are returned, along with the term papers with the comments of the rater. If PEERRATE is used near the end of the semester, sufficient time must be allowed for the students to rate the papers, return them, and then for the instructor to run the program and return the printouts. A week is usually sufficient time for rating, computer run, and the return of the papers and printouts.

Aside from keypunching the ratings, a PEERRATE user must prepare a problem card which controls program parameters and options, plus a list of student names that will appear on the printout. Table 1 presents Benson's outline of the problem card. Notice that the problem card would contain the mean and standard deviation that the instructor selects for his class.

RESULTS FROM PEERRATE AND STUDENT REACTIONS

If students are properly briefed on the rating task and if the class as a whole has gained some knowledge of the subject matter, I have found the adjusted class ratings to be a good estimate of the rank orders I placed on the papers when reading them myself. PEERRATE routinely prints a matrix of inter-correlations of all the variables in the system. The correlations obtained between paper grades and the writer's ability as a rater ranged between $r=.632$ and $r=.703$ ($P<.01$). This is suggestive of internal validity of the technique since students who write good papers show up as better paper raters as well.

Table 2 presents a sample of the printout the instructor receives reflecting the rank orders of the entire class by paper grade (performance score), quality of rating for each paper, and, finally, the rater's ability as rater. This table serves as an important diagnostic for the teacher before the return of the student printouts. If the instructor chooses not to read all the papers, he can still supervise the accuracy of the performance scores (paper grades) by reading a subset of papers made up of: (a) papers with low grades; (b) papers with high grades; and (c) papers with low rater quality scores, especially if they

coincide with low performance grades. Low rating quality scores (the middle set of columns in Table 2) indicate rater disagreement about the paper after the calculation of adjusted class ratings. Examples of papers that would be checked in Table 2 would be ID numbers 9, 28, 20, 36, 19 (papers with low scores and low rater agreement); number 22 (high score, low agreement); and 39, 15, and 16 (high scores). Should the instructor disagree with the paper's grade, he simply overrides it -- either on the student's printout or by indicating the change on a subsequent run of the program.

A second level of pedagogical diagnosis occurs in the student's printed report. A sample is included in Table 3. For this student, the overall paper grade is 87 (upper right hand corner). The five criteria used in grading are listed next, with their respective norm adjusted scores: content, organization, writing style, originality, and overall. Notice that the pattern of these scores can indicate to a student where the relative strengths and weaknesses of his paper lie. The student in Table 3 has a good score for content and originality but his organization and writing style (grammar) suffered in comparison. The end column in this section of Table 3 lists the weight given each of the criteria in the computation of the adjusted class rating (the 87 in this case).

The student can also see the adjusted ratings given his paper by each rater on each of the criteria. These are displayed in the middle of Table 3 "Report of Ratings Received from Individual Raters". The first row of numbers from each rater (item ratings 1 through 5) contains the ratings on content, organization, etc. The next three rows contain measures of that rater's bias.

High negative numbers indicate a stiff grader; high positives a lenient one. The closer to zero the numbers appear, the closer the rater came to the class rating characteristics. "Rater residual" refers to the standard deviation of the residual error remaining in the rater's ratings after his bias was adjusted by the program.

The final section of Table 3 displays measures of this student's rating ability. The top score "Overall Score for Rating Performance" is a composite measure made up from the next four measures of rating agreement (defined earlier). The instructor sets not only the mean and deviation of the student's paper grades, but also the mean and deviation of the distribution of student scores for rater performance. As mentioned earlier, the score for rating can then also become a part of the student's grade on the assignment as a whole. In the class example used here, the mean was set at 85 and the deviation at 7, thus 2/3 of the class scores fell between 78 and 92. The paper assignment was worth 50 points (not 50%) in the grading scheme for the course, so the students could calculate their point score on the paper by dividing their overall performance score in half. PEERRATE can print results on any point scale and transform one scale into another. The overall score could easily have been printed on a 0 to 50 scale.

For the past four semesters, an anonymous questionnaire has been distributed following the peer rating exercise. The question form solicited comments from student users and also asked specific questions about degree of interest and the amount of learning the students felt PEERRATE added to the course. Table 4 presents the results of the survey. The responses to the last two questions in Table 4 suggest that the majority of students feel they learn

more by reading each other's papers and an overwhelming majority (79%) feel that PEERRATE is at least worth further development in conjunction with the course taken.

Was the questionnaire a sensitive instrument or did it suffer from student social compliance with a project initiated by the instructor? Further analysis of the classes making up the survey data showed that one could divide them into groups that had filled out the questionnaire after rating but before grade feed-back and those who responded after rating and receiving printed results. The percentage of responses "yes" to the "continue" question was unchanged. However, the percentage in the other two categories shifted significantly ($\chi^2 = 18.015$, $df = 2$, $P < .01$), with more "no's" in one post-feed-back group. It is to be expected that some students will always be dissatisfied with any grading system and certainly PEERRATE is not a grading panacea. The data indicate that the students did react honestly. The majority response to the "learn more" question alone suggests that the system has didactic as well as pragmatic value.

Students tend to react more positively to the system if they feel the instructor is doing his "share" of the work and if they can have input in the form of setting initial criteria and including a "self" rating along with other raters. The difference between self ratings and the group adjusted rating becomes another measure of rating skill. In one class of 77 students, the raw score mean rating for the class and the raw score mean self rating differed by about two points, a difference that was non-significant by t-test.

When the PEERRATE system is first announced in a class, it sometimes raises questions. Some students initially feel they are not competent or well enough informed to rate each other. This is especially true in classes like child psychology which attract students from a variety of majors and class years. While PEERRATE might be most suited to homogenious upper class courses in a given major field, the instructor can structure class assignments so that it is amenable to more heterogenious classes. One strategy that has worked at Rutgers is to assign a term paper that is based on a set of readings that is common to the class. For example, I have assigned a book of secondary source readings and required term papers that traced the primary source materials underlying a chosen subset of the readings.

Experiments assessing the objective amount of learning increase attributable to PEERRATE (rather than student opinion) are in the planning stage. The peer rating system described here, however, would seem to have benefits beyond learning the content in a given course. The experiences of "sitting on the other side of the professor's desk", of participating in the definition and analysis of rating criteria, and of constructing a paper with multiple readers in mind assist in developing the kind of independent, critical thought processes that are the goals of a liberal education.

PEERRATE AVAILABILITY

PEERRATE is written in FORTRAN. It consists of over 4,000 input cards, many of which are comments and diagnostics to aid the user. The program is

currently mounted on the New Jersey State Educational Computer Network (IBM 370) and is available to the public. It can be run on any medium-sized computer. Storage requirement is about 240K. The typical user in higher education need only keypunch his system's JCL cards, the PEERRATE problem cards, and his student's input ratings.

REFERENCES

Benson, Purnell H. "A Computerized System of Student Grading of Student Assignments." Third Annual New Jersey Conference on the Use of Computers in Higher Education. New Brunswick, New Jersey: New Jersey Educational Computer Network, Inc., 1976, 21-30.

Benson, Purnell H. A Supervised Student Grading System. American Statistical Association, Social Statistics Section, 1977 (in press).

Benson, Purnell H. Documentation for the Use of PEERRATE. Rutgers University Center for Computing and Information Services, Newark, New Jersey, 07102.

Mosteller, Frederick. Remarks on the Method of Paired Comparisons: The Least Squares Solution Assuming Equal Standard Deviations and Equal Correlations. Psychometrika, 16, 1951, 3-9.

TABLE 1

INPUT PARAMETERS FOR PROBLEM CARD AND OTHER CONTROL CARDS

1. Problem card.

Parameter
Highest ID No. used for student in class
Number of tasks rated
Number of sets of records
Exponent for skewing rater scores
Grade norm imposed by instructor
Grade stretch imposed by instructor
Section number of class
Multiplier for deviation in rater score
Multiplier for variability in rater score
Origin for deviation in rater score
Origin for variability in rater score
Minimum number of ratings for inclusion
Minimum permissible grade
Maximum permissible grade
Enter 1 if instructor grade imposed to override class rating
Enter 1 if grade is average of tasks, not sum
Enter 1 to print matrixes, 2 for solution, 3 for both
Enter 1 for printout of individual grade sheets
Number of iterations if rater weights calculated

2. Point spread for tasks (exercises or assignment parts).

3. Weights for tasks.

4. Rating norms and rating spreads for tasks.

5. Weights for raters, or parameters for iterating weights.

6. Number of student ratings inputted by rater in each record.

7. Names of tasks.

8. List of names in class, ID Nos., and 1 punch if instructor grade imposed.

9. Rating records supplied by students, and by instructor if acting as a student.

10. Instructor's imposed grade and grade weight. (optional)

For column assignments for input 2 through 10, consult the particular version of program used.

TABLE 2

26A. PERF SCORES, RATING QUALITIES, AND RATER ABILITIES IN ASCENDING ORDER

PERF SCORE			RATING QUALITY			RATER ABILITY		
ID NO.	SUB NO.	SCORE	ID NO.	SUB NO.	SCORE	ID NO.	SUB NO.	SCORE
9	9	63	9	9	67	19	19	*****
28	28	64	36	36	68	21	21	*****
20	20	70	12	12	76	36	36	68
36	36	71	28	28	78	20	20	72
19	19	74	20	20	79	33	33	73
27	27	74	7	7	79	9	9	74
41	41	74	19	19	79	28	28	76
3	3	76	26	26	81	38	38	76
12	12	77	3	3	81	3	3	78
24	24	77	21	21	81	12	12	79
13	13	78	24	24	82	7	7	82
23	23	80	22	22	82	24	24	82
37	37	81	10	10	83	41	41	82
40	40	81	37	37	83	16	16	85
33	33	81	41	41	83	6	6	85
31	31	82	33	33	84	13	13	86
7	7	82	31	31	84	37	37	86
4	4	83	39	39	85	26	26	86
32	32	83	38	38	86	34	34	86
14	14	83	40	40	87	22	22	87
34	34	83	23	23	87	32	32	87
26	26	84	11	11	89	17	17	87
11	11	85	6	6	89	11	11	87
38	38	85	13	13	89	4	4	88
17	17	86	16	16	89	29	29	88
5	5	87	17	17	89	23	23	88
21	21	87	34	34	89	10	10	89
8	8	87	15	15	90	2	2	89
6	6	87	8	8	90	39	39	89
30	30	88	14	14	90	14	14	89
18	18	88	32	32	90	5	5	90
2	2	90	2	2	91	27	27	90
1	1	90	5	5	91	15	15	90
35	35	90	35	35	91	31	31	91
29	29	90	29	29	92	30	30	92
10	10	91	27	27	93	25	25	92
25	25	92	4	4	94	8	8	93
16	16	93	30	30	95	40	40	94
15	5	94	18	18	95	35	35	94
22	22	95	25	25	97	18	18	95
39	39	97	1	1	99	1	1	95

26. PROJECT 1277 NUMBER 1 PROF. ANTHONY LUTKUS DECEMBER 1977
INDIVIDUAL SCORE SHEETS FOLLOW. FOR OUTPUT ON CARDS, TAPE OR DISK, USE A SINGL
SKIP NAME. USE 4 INTEGERS FOR IDENTIFICATIONS, THEN RATINGS AS DECIMALS.
UNIT NO. 26 OUTPUTS NAME, STUDY NO., SECTION NO., RATEE ID NO., RATER ID NO.,
THIS OUTPUT PRECEDES OUTPUT OF INDIVIDUAL SCORE SHEETS ON CARDS, TAPE OR DISK.

TABLE 3

A Sample Student Printout

CALLARI.J.J. ID. NO. 6 PROJECT 1277 NUMBER 1 ***PEER-RATIN

PROF. ANTHONY LUTKUS DECEMBER 1977 #8

**** OVERALL SCORE FOR PERFORMANCE OF ASSIGNMENT(S)---------- 87
THIS IS SCORE RECEIVED ON ASSIGNMENT WITH 3 RATERS RECORDS

POINTS FOR ITEMS ADJUSTED TO NORMS FOR GROUP:

ITEM NAME	POINT BASIS	ITEM RATING ADJ TO NORM	WEIGHT FOR ITEM
CONTENT	100	91	0.210
ORGANIZATION	100	80	0.180
WRITING STYLE	100	84	0.150
ORIGINALITY	100	90	0.060
OVER-ALL	100	89	0.400

SUCCESSIVE ITEM SELF-RATINGS BY PEER ON POINT BASIS ABOVE:
90 85 95 95 95

****REPORT OF RATINGS RECEIVED FROM INDIVIDUAL RATERS ON POINT BASIS ABOVE

RATER ID NO. WT.	COMPARED WITH ADJ. GROUP RTG	SCORE GIVEN	ITEM RATING GIVEN BY RATER 1	2	3	4	5
1 RTGS FROM RTR:		80	85	80	70	90	80
	RATER BIAS	2.9	-4.6	0.8	-8.5	0.3	-6.9
	RATER RANGE DIF	2.1	-0.7	3.7	5.5	-0.8	-0.5
	RATER RESIDUAL	2.6	2.5	8.9	14.7	3.4	3.2
1 RTGS FROM RTR:		92	90	85	95	95	95
	RATER BIAS	3.1	3.1	7.3	7.0	5.2	6.6
	RATER RANGE DIF	2.2	-2.9	-0.8	-2.9	-1.2	-2.6
	RATER RESIDUAL	2.1	5.9	7.2	6.9	4.2	5.1
1 RTGS FROM RTR:		96	100	95	95	95	95
	RATER BIAS	3.1	6.3	7.6	7.8	6.2	4.4
	RATER RANGE DIF	2.2	0.7	-0.9	5.4	0.1	1.3
	RATER RESIDUAL	2.1	4.7	7.7	10.7	2.0	3.0

***** OVERALL SCORE FOR RATING PERFORMANCE---------- 92
THIS IS SCORE RECEIVED FOR RATING PERFORMANCE WITH 4 RATEES

(CALCULATIONS WHICH FOLLOW ARE DONE ON BASIS OF 100 POINTS)
SCORE FOR DEVIATION FROM ADJ GROUP RATINGS---------- 87
WEIGHT FOR DEV FROM ADJ GROUP RATINGS---------- 1.0

SCORE FOR DIFFERENCE IN RANGE FROM ADJ GROUP RATINGS------- 93
WEIGHT, DIF IN RANGE FROM ADJ GROUP RATINGS------ 1.0

SCORE FOR RESIDUAL ERROR FROM ADJ GROUP RATINGS---------- 100
WEIGHT, RESIDUAL ERROR FROM ADJ GROUP RATINGS---- 1.0

SCORE FOR ERROR IN SELF-RATINGS FROM ADJ GROUP RATINGS----- 98
WEIGHT, ERROR IN SELF-RTGS FROM ADJ GROUP RTGS--- 1.0
(IF BLANK, NO EVALUATION OF SELF-RATINGS IS MADE)

TABLE 4
RESULTS OF PEERRATE FOLLOW-UP SURVEY (N=186)

Question		Percent Of Respondents
Clarity of instructions?	Extremely clear	30
	Quite clear	59
	Somewhat clear	10
	Not at all clear	1
PEERRATE make the course interesting?	More	55
	Less	8
	Little difference	37
Amount of learning?	More	64
	Less	4
	Same	32
Continue PEERRATE in this course?	Yes	41
	No	21
	Worth further trial	38

APPENDIX A

NAME 1 , YOUR PAPER # 8A

PRINCIPLES OF PSYCHOLOGY
Dr. Anthony D. Lutkus

PEER-RATING SHEET

****

SUGGESTED RATING SCALE

Outrageous	Poor	Adequate	Good	Excellent
1-2	3-4	5-6	7-8	9-10

****

1st Paper # 11A

Content 7 Organization 6 Writing Style 8 Originality 7
Over-All 7

2nd Paper # 25A

Content 6 Organization 9 Writing Style 9 Originality 7
Over-All 7

3rd Paper # 39A

Content 10 Organization 11 Writing Style 9 Originality 7
Over-All 9 11

Rate Your Own Paper # 8A

Content 9 Organization 9 Writing Style 7 Originality 9
Over-All 9

11/77

APPENDIX B

PEER RATING DATA SHEET

*** RETURN RATING SHEET TO CLASS OR TO DEAN'S OFFICE, 325 HILL HALL ***
BY THURSDAY, DECEMBER 8

You will be rating your fellow students' papers on five categories that were discussed in class. You should give your rating on each dimension as independently as possible. The rating score can range from 1 to 10 points according to the guideline below. The ratings will remain annonomous. Remember that the computer program is concerned only with the intervals between the ratings. There is also a provision to adjust the mean of the distribution and the relative dispersion around that mean. Therefore you will not benefit yourself or others by giving falsely high ratings to every one. Try to use the whole scale and to assess the papers as fairly as possible. Record your ratings in the spaces on the rating sheet.

SUGGESTED RATING SCALE

Outrageous	Poor	Adequate	Good	Excellent
1-2	3-4	5-6	7-8	9-10

THE RATING CATEGORIES

CONTENT: Includes adequacy of the summary of the articles within the page limitations. Are outside sources brought in? That is, the text, lectures or other books, articles from journals. Are several points of view given if the topic calls for them?

ORGANIZATION: Is there some logical ordering to the presentation of the material? If so, can the reader easily uncover the structure? That is, are there outlines, headings, introductory and/or summary statements? Do paragraphs join together coherently? Are different issues, arguments and the data to support them clearly delineated? Does the author build a case using the material he has presented or does he/she merely summarize another author?

WRITING STYLE: Is the paper in literate English? Do subject and verbs agree in number and tense? Are the sentences complete? Is the writing style clear, i.e., not rambling and/or awkardly constructed? Is it a good "composition"?

ORIGINALITY: Does the paper sound suspiciously like a text book? Or does it sound like a student writer? Honest student attempts (with their occassional errors and awkardness) are to be preferred to prose that is too closely paraphrased from published material. Does the writer himself have an opinion? Is it based on the material he has presented?

OVER-ALL: Give your best estimate of the paper's over-all quality. Do not, however, simply take an average of the four categories above.